

What Is Claimed Is:

1. A method for generating an acoustic fingerprint of a digital audio signal, comprising:

downsampling a received digital audio signal based upon a predetermined frequency;

subdividing the downsampled, digital audio signal into a beginning portion, a middle portion and an end portion;

extracting a plurality of beginning frames, a plurality of middle frames and a plurality of end frames from the beginning, middle and end portions of the downsampled, digital audio signal, respectively, each frame having a predetermined number of samples;

generating a plurality of frame vectors from the plurality of beginning, middle and end frames, each frame vector including a plurality of acoustic features;

creating an acoustic fingerprint of the digital audio signal based on the plurality of frame vectors; and

storing the acoustic fingerprint in a database.

2. The method according to Claim 1, wherein said generating a frame vector for each frame includes:

computing a plurality of time domain features from the predetermined number of samples within the frame;

computing a plurality of spectral domain features from the predetermined number of samples within the frame;

computing a plurality of wavelet domain features from the predetermined number of samples;

computing a plurality of second stage spectral features from the predetermined spectral domain FFT results; and

creating the frame vector.

3. The method according to Claim 2, wherein said generating a frame vector for each frame includes:

Applying a logarithmic conversion to the plurality of spectral power bands;

Creating an indexed array based on the plurality of log-converted spectral power bands;

Determining a number of beats within the indexed array; and

Including the number of beats within the frame vector.

4. The method according to Claim 2, wherein the wavelet domain features include using a Haar wavelet transform, using a Blackman-Harris window.

5. The method according to Claim 1, further comprising:

Downmixing the downsampled audio signal to create a single channel, downsampled digital audio signal.

6. The method according to Claim 2, wherein the predetermined frequency is about 11025 hz.

7. The method according to Claim 2, wherein:

The predetermined number of samples is about 96,000;

The plurality of beginning frames includes five frames;

The plurality of middle frames includes three frames; and

The plurality of end frames includes five frames.

8. The method according to claim 1, wherein said extracting a plurality of middle frames includes:

Determining a total number of frames within the plurality of middle frames;

Calculating a number of frames to average by dividing the total number of frames by a constant; and

Averaging the plurality of middle frames, based on the number of frames to average, to create the constant number of frames.

9. The method according to Claim 1, wherein the plurality of time domain features include a zero crossing rate, a zero crossing mean, a sample mean and RMS ratio, a mean energy value, and a mean energy delta value.

10. A method for generating an acoustic fingerprint frame vector from a frame extracted from a digital audio signal, comprising:

Computing a plurality of time domain features from a plurality of samples within the frame;

Applying a window function to the plurality of samples;

Applying a Fast Fourier Transform to the plurality of windowed samples to create a plurality of spectral power bands;

Determining the number of beats from the spectral power bands;

Selecting one or more output spectral power bands and using one or more first stage FFT outputs as input for a second Fast Fourier Transform;

Selecting one or more output second stage power bands, summing across all output second stage Fast Fourier Transforms, and normalizing the resulting sum by the number of input Transforms;

Creating an acoustic fingerprint frame vector including the plurality of second stage normalized bands, the plurality of time domain features and the number of beats; and

Storing the acoustic fingerprint frame vector in a memory.

11. The method according to Claim 10, wherein the plurality of time domain features include a zero crossing rate, a zero crossing mean, a sample mean and RMS ratio, a mean energy value, and a mean energy delta value.

12. The method according to claim 10, wherein the wavelet domain features include using a Haar wavelet transform, using a Blackman-Harris window.

13. The method according to Claim 10, wherein the plurality of samples consists of about 96,000 samples.

14. An information storage medium storing information operable to perform the method of any of the preceding claims.

15. A system as substantially herein described.

16. A system for generating an acoustic fingerprint of a digital audio signal, comprising:

means for downsampling a received digital audio signal based upon a predetermined frequency;

means for subdividing the downsampled, digital audio signal into a beginning portion, a middle portion and an end portion;

means for extracting a plurality of beginning frames, a plurality of middle frames and a plurality of end frames from the beginning, middle and end portions of the downsampled, digital audio signal, respectively, each frame having a predetermined number of samples;

means for generating a plurality of frame vectors from the plurality of beginning, middle and end frames, each frame vector including a plurality of spectral residual bands and a plurality of time domain features;

means for creating an acoustic fingerprint of the digital audio signal based on the plurality of frame vectors; and

means for storing the acoustic fingerprint in a database.

17. The system according to Claim 16, wherein said means for generating a frame vector for each frame includes:

means for computing a plurality of time domain features from a plurality of samples within the frame;

means for applying a window function to the plurality of samples;

means for applying a Fast Fourier Transform to the plurality of windowed samples to create a plurality of spectral power bands;

means for determining the number of beats from the spectral power bands;

means for selecting one or more output spectral power bands, and using one or more first stage FFT outputs as input for a second Fast Fourier Transform;

means for selecting one or more output second stage power bands, summing across all output second stage Fast Fourier Transforms, and normalizing the resulting sum by the number of input Transforms;

means for creating an acoustic fingerprint frame vector including the plurality of second stage normalized bands, the plurality of time domain features and the number of beats; and

means for creating the frame vector.

18. The system according to Claim 17, wherein said means for generating a frame vector for each frame includes:

means for applying a logarithmic conversion to the plurality of spectral power bands;

means for creating an indexed array based on the plurality of log-converted spectral power bands;

means for determining a number of beats within the indexed array; and

means for including the number of beats within the frame vector.

19. The system according to Claim 17, wherein the wavelet domain features include using a Haar wavelet transform, using a Blackman-Harris window.

20. The system according to Claim 17, wherein the predetermined number of samples consists of about 96,000 samples.

21. A method of sequencing digital media playback, comprising:

receiving a plurality of acoustic fingerprints as the seed;

selecting a weight bank for comparing the seed acoustic fingerprints;

comparing the seed fingerprint with a plurality of reference fingerprints using a selected weight bank;

selecting a subset of the reference fingerprints based on their similarity with the seed fingerprint;

applying a sort mechanism to the resultant subset; and

sequencing digital media playback using resultant sorted subset.

22. The method according to Claim 21, wherein said selecting a weight bank includes:

comparing the seed fingerprint with a plurality of weight class reference vectors;

and

selecting the weight class vector which is most similar to the seed fingerprint.

23. The method according to claim 21, wherein applying a sort mechanism includes:

randomly selecting a start acoustic fingerprint from the result set and moving it to the final sorted set;

computing the similarity between the last acoustic fingerprint in the sorted set and each remaining acoustic fingerprint in the result set;

moving the acoustic fingerprint with the highest similarity into the final sorted set; and

repeating until all acoustic fingerprints have been moved into the final sorted set.

24. The method according to claim 21, wherein applying a sort mechanism includes:

randomly selecting an acoustic fingerprint from the result set and moving it to the final sorted set; and

repeating until all acoustic fingerprints have been moved into the final sorted set.

25. The method according to claim 21, wherein sequencing digital media playback includes:

mapping each result acoustic fingerprint to a media identifier;

mapping each media identifier to a digital media element; and

generating a playlist containing the sorted digital media elements.

26. The method according to Claim 21, wherein selecting a weight bank additionally adds the means to retrain a weight bank which includes:

providing a display component wherein a plurality of sliders elements are linked to one or more features within the selected weight bank.

27. The method according to Claim 21, wherein selecting a weight bank additionally adds the means to retrain a weight bank which includes:

providing a user interface to allow a plurality of fingerprints to be marked as more similar;

comparing said plurality of fingerprints, and raising the weight of similar features by a scaling factor, and reducing the weight of dissimilar features by said scaling factor; and

normalizing the modified weights by said scaling factor.

28. The method according to Claim 21, wherein selecting a weight bank additionally adds the means to retrain a weight bank which includes:

Providing a user interface to allow a plurality of fingerprints to be marked as less similar;

Comparing said plurality of fingerprints, and lowering the weight of similar features by a scaling factor, and raising the weight of dissimilar features by said scaling factor; and

Normalizing the modified weights by said scaling factor.

29. The method according to claim 21, wherein receiving a plurality of acoustic fingerprints as seed includes:

Generating an identification acoustic fingerprint from an input digital audio source;

Resolving the identification acoustic fingerprint using a reference acoustic fingerprint database to return a sequencing acoustic fingerprint identifier; and

Retrieving a reference sequencing acoustic fingerprint from a reference database using said sequencing acoustic fingerprint identifier.